18.175 Theory of Probability
Fall 2008

Section 4

Laws of Large Numbers.

Consider a r.v. X and sequence of r.v.s $(X_n)_{n\geq 1}$ on some probability space. We say that X_n converges to X *in probability* if for all $\varepsilon > 0$
$$\lim_{n\to\infty} \mathbb{P}(|X_n - X| \geq \varepsilon) = 0.$$
We say that X_n converges to X *almost surely* or *with probability* 1 if
$$\mathbb{P}(\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)) = 1.$$

Lemma 9 *(Chebyshev's inequality) If a r.v. $X \geq 0$ then for $t > 0$,*
$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}X}{t}.$$

Proof.
$$\mathbb{E}X = \mathbb{E}X\mathrm{I}(X < t) + \mathbb{E}X\mathrm{I}(X \geq t) \geq \mathbb{E}X\mathrm{I}(X \geq t) \geq t\mathbb{E}\mathrm{I}(X \geq t) = t\mathbb{P}(X \geq t).$$

□

Theorem 4 *(Weak law of large numbers) Consider a sequence of r.v.s $(X_i)_{i\geq 1}$ that are centered, $\mathbb{E}X_i = 0$, have finite second moments, $\mathbb{E}X_i^2 \leq K < \infty$ and are uncorrelated, $\mathbb{E}X_iX_j = 0, i \neq j$. Then*
$$S_n = \frac{1}{n}\sum_{i\leq n} X_i \to 0$$
in probability.

Proof. By Chebyshev's inequality we have
$$\begin{aligned} \mathbb{P}(|S_n - 0| \geq \varepsilon) &= \mathbb{P}(S_n^2 \geq \varepsilon^2) \leq \frac{\mathbb{E}S_n^2}{\varepsilon^2} \\ &= \frac{1}{n^2\varepsilon^2}\mathbb{E}(X_1 + \cdots + X_n)^2 = \frac{1}{n^2\varepsilon^2}\sum_{i\leq n}\mathbb{E}X_i^2 \leq \frac{nK}{n^2\varepsilon^2} = \frac{K}{n\varepsilon^2} \stackrel{n\to\infty}{\longrightarrow} 0. \end{aligned}$$

□

Example. Before we turn to a.s. convergence results, let us note that convergence in probability is weaker than a.s. convergence. For example, consider a probability space which is a circle of circumference 1 with uniform measure on it. Consider a sequence of r.v. on this probability space defined by
$$X_k(x) = \mathrm{I}\Big(x \in \Big[1 + \frac{1}{2} + \cdots + \frac{1}{k}, 1 + \cdots + \frac{1}{k+1}\Big) \text{ mod } 1\Big).$$

Then $X_k \to 0$ in probability, since for $0 < \varepsilon < 1$

$$\mathbb{P}(|X_k - 0| \geq \varepsilon) = \frac{1}{k+1} \to 0$$

but, clearly, X_k does not converge a.s. because the series $\sum_{k\geq 1} 1/k$ diverges and, as a result, each point x on the sphere will fall into the above intervals infinitely many times, i.e. it will satisfy $X_k(x) = 1$ for infinitely many k.

$\square$

Lemma 10 *Consider a sequence $(p_i)_{i\geq 1}$ such that $p_i \in [0,1)$. Then*

$$\prod_{i\geq 1}(1-p_i) = 0 \Longleftrightarrow \sum_{i\leq 1} p_i = +\infty.$$

Proof. "$\Longleftarrow$". Using that $1 - p \leq e^{-p}$ we get

$$\prod_{i\leq n}(1-p_i) \leq \exp(-\sum_{i\leq n} p_i) \to 0 \text{ as } n \to \infty.$$

"$\Longrightarrow$". We can assume that $p_i \leq \frac{1}{2}$ for $i \geq m$ for large enough m because, otherwise, the series obviously diverges. Since $1 - p \geq e^{-2p}$ for $p \leq 1/2$ we have

$$\prod_{m\leq i\leq n}(1-p_i) \geq \exp\Big(-2\sum_{m\leq i\leq n} p_i\Big)$$

and the result follows.

$\square$

Lemma 11 *(Borel-Cantelli) Consider a sequence $(A_n)_{n\geq 1}$ of events $A_n \in \mathcal{A}$ on probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Consider an event*

$$A_n \quad i.o. := \limsup A_n := \bigcap_{n\geq 1}\bigcup_{m\geq n} A_m$$

that A_n occur infinitely often. Then

1. $\sum_{n\geq 1} \mathbb{P}(A_n) < \infty \Longrightarrow \mathbb{P}(A_n \ i.o.) = 0$.
2. *If A_n are independent then* $\sum_{n\geq 1} \mathbb{P}(A_n) = +\infty \Longrightarrow \mathbb{P}(A_n \ i.o.) = 1$.

Proof. 1. If $B_n = \bigcup_{m\geq n} A_m$ then $B_n \supseteq B_{n+1}$ and by continuity of measure

$$\mathbb{P}(A_n \text{ i.o}) = \mathbb{P}\Big(\bigcap_{n\geq 1} B_n\Big) = \lim_{n\to\infty} \mathbb{P}(B_n).$$

We have

$$\mathbb{P}(B_n) = \mathbb{P}\Big(\bigcup_{m\geq n} A_m\Big) \leq \sum_{m\geq n} \mathbb{P}(A_m) \to 0 \text{ as } n \to +\infty \text{ because } \sum_{m\geq 1} \mathbb{P}(A_m) < \infty.$$

2. We have

$$\begin{aligned}\mathbb{P}\Big(\Omega \setminus B_n\Big) &= \mathbb{P}\Big(\Omega \setminus \bigcup_{m\geq n} A_m\Big) = \mathbb{P}\Big(\bigcap_{m\geq n} (\Omega \setminus A_m)\Big) \\ &= \prod_{m\geq n} \mathbb{P}(\Omega \setminus A_m)(\text{by independence}) = \prod_{m\geq n}(1 - \mathbb{P}(A_m)) = 0,\end{aligned}$$

by Lemma 10, since $\sum_{m\geq n} \mathbb{P}(A_m) = +\infty$. Therefore, $\mathbb{P}(B_n) = 1$ and $\mathbb{P}(A_n \text{ i.o}) = \mathbb{P}\Big(\bigcap_{n\geq 1} B_n\Big) = 1$.

$\square$

Strong law of large numbers. The following simple observation will be useful. If a random variable $X \geq 0$ then $\mathbb{E}X = \int_0^\infty \mathbb{P}(X \geq x)dx$. Indeed,

$$\mathbb{E}X = \int_0^\infty x dF(x) = \int_0^\infty \int_0^x 1 ds dF(x) = \int_0^\infty \int_s^\infty 1 dF(x) ds = \int_0^\infty \mathbb{P}(X \geq s) ds.$$

For $X \geq 0$ such that $\mathbb{E}X < \infty$ this implies

$$\sum_{i \geq 1} \mathbb{P}(X \geq i) \leq \int_0^\infty \mathbb{P}(X \geq s) ds = \mathbb{E}X < \infty.$$

Theorem 5 *(Strong law of large numbers) If $\mathbb{E}|X| < \infty$ and $(X_i)_{i \geq 1}$ are i.i.d. copies of X then*

$$\frac{1}{n} \sum_{i=1}^n X_i \to \mathbb{E}X_1 \textit{ almost surely (a.s.).}$$

Proof. The proof will proceed in several steps.

1. First, without loss of generality we can assume that $X_i \geq 0$. Indeed, for signed r.v.s we can decompose $X_i = X_i^+ - X_i^-$ where

$$X_i^+ = X_i \mathrm{I}(X_i \geq 0) \text{ and } X_i^- = |X_i| \mathrm{I}(X_i < 0)$$

and the general result would follow since

$$\frac{1}{n} \sum_{i=1}^n X_i = \frac{1}{n} \sum_{i=1}^n X_i^+ - \frac{1}{n} \sum_{i=1}^n X_i^- \to \mathbb{E}X_1^+ - \mathbb{E}X_1^- = \mathbb{E}X_1.$$

Thus, from now on we assume that $X_i \geq 0$.

2. (Truncation) Next, we can replace X_i by $Y_i = X_i \mathrm{I}(X_i \leq i)$ using Borel-Cantelli lemma. We have

$$\sum_{i \geq 1} \mathbb{P}(X_i \neq Y_i) = \sum_{i \geq 1} \mathbb{P}(X_i > i) \leq \mathbb{E}X_1 < \infty$$

and Borel-Cantelli lemma implies that $\mathbb{P}(\{X_i \neq Y_i\} \text{ i.o.}) = 0$. This means that for some (random) i_0 and for $i \geq i_0$ we have $X_i = Y_i$ and, therefore,

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^n X_i = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^n Y_i.$$

It remains to show that if $T_n = \sum_{i=1}^n Y_i$ then $\frac{T_n}{n} \to \mathbb{E}X$ a.s.

3. (Limit over subsequences) We will first prove this along the subsequences $n(k) = \lfloor \alpha^k \rfloor$ for $\alpha > 1$. For any $\varepsilon > 0$,

$$\begin{aligned} \sum_{k \geq 1} \mathbb{P}\Big(|T_{n(k)} - \mathbb{E}T_{n(k)}| \geq \varepsilon n(k)\Big) &\leq \sum_{k \geq 1} \frac{1}{\varepsilon^2 n(k)^2} \mathrm{Var}(T_{n(k)}) = \sum_{k \geq 1} \frac{1}{\varepsilon^2 n(k)^2} \sum_{i \leq n(k)} \mathrm{Var}(Y_i) \\ &\leq \sum_{k \geq 1} \frac{1}{\varepsilon^2 n(k)^2} \sum_{i \leq n(k)} \mathbb{E}Y_i^2 = \frac{1}{\varepsilon^2} \sum_{i \geq 1} \mathbb{E}Y_i^2 \sum_{k : n(k) \geq i} \frac{1}{n(k)^2} \\ &\overset{(*)}{\leq} K \sum_{i \geq 1} \mathbb{E}Y_i^2 \frac{1}{i^2} = K \sum_{i \geq 1} \frac{1}{i^2} \int_0^i x^2 dF(x) \leq K \times (**) \end{aligned}$$

where $F(x)$ is the law of X and a constant $K = K(\alpha)$ depends only on α. (*) follows from

$$\frac{\alpha^k}{2} \leq n(k) = \lfloor \alpha^k \rfloor \leq \alpha^k$$

and if $k_0 = \min\{k : \alpha^k \geq i\}$ then

$$\sum_{n(k)\geq i} \frac{1}{n(k)^2} \leq \sum_{\alpha^k \geq i} \frac{4}{\alpha^{2k}} = \frac{4}{\alpha^{2k_0}(1-\frac{1}{\alpha^2})} \leq \frac{K}{i^2}.$$

We can continue,

$$\begin{aligned}
(**) &= \sum_{i\geq 1} \frac{1}{i^2} \sum_{m<i} \int_m^{m+1} x^2 dF(x) = \sum_{m\geq 0} \sum_{i>m} \frac{1}{i^2} \int_m^{m+1} x^2 dF(x) \\
&\leq \sum_{m\geq 0} \frac{1}{m+1} \int_m^{m+1} x^2 dF(x) \leq \sum_{m\geq 0} \int_m^{m+1} x dF(x) = \mathbb{E}X < \infty.
\end{aligned}$$

Thus, we proved that

$$\sum_{k\geq 1} \mathbb{P}\Big(|T_{n(k)} - \mathbb{E}T_{n(k)}| \geq \varepsilon n(k)\Big) < \infty$$

and Borel-Cantelli lemma implies that

$$\mathbb{P}\Big(|T_{n(k)} - \mathbb{E}T_{n(k)}| \geq \varepsilon n(k) \text{ i.o.}\Big) = 0.$$

This means that for some (random) k_0

$$\mathbb{P}\Big(\forall k \geq k_0, |T_{n(k)} - \mathbb{E}T_{n(k)}| \leq \varepsilon n(k)\Big) = 1.$$

If we take a sequence $\varepsilon_m = \frac{1}{m}$, this implies that

$$\mathbb{P}\Big(\forall m \geq 1, k \geq k_0(m), |T_{n(k)} - \mathbb{E}T_{n(k)}| \leq \frac{1}{m} n(k)\Big) = 1$$

and this proves that

$$\frac{T_{n(k)}}{n(k)} - \frac{\mathbb{E}T_{n(k)}}{n(k)} \to 0 \text{ a.s.}$$

On the other hand,

$$\frac{1}{n(k)} \mathbb{E}T_{n(k)} = \frac{1}{n(k)} \sum_{i\leq n(k)} \mathbb{E}X_i \mathrm{I}(X_i \leq i) \to \mathbb{E}X \text{ as } k \to \infty,$$

by Lebesgue's dominated convergence theorem. We proved that

$$\frac{T_{n(k)}}{n(k)} \to \mathbb{E}X \text{ a.s.}$$

4. Finally, for j such that

$$n(k) \leq j < n(k+1) = n(k)\frac{n(k+1)}{n(k)} \leq n(k)\alpha^2$$

we can write

$$\frac{1}{\alpha^2} \frac{T_{n(k)}}{n(k)} \leq \frac{T_j}{j} \leq \alpha^2 \frac{T_{n(k+1)}}{n(k+1)}$$

and, therefore,

$$\frac{1}{\alpha^2} \mathbb{E}X \leq \liminf \frac{T_j}{j} \leq \limsup \frac{T_j}{j} \leq \alpha^2 \mathbb{E}X \text{ a.s.}$$

Taking $\alpha = 1 + m^{-1}$ and letting $m \to \infty$ proves that $\lim_{j\to\infty} \frac{T_j}{j} = \mathbb{E}X$ a.s.

$\square$